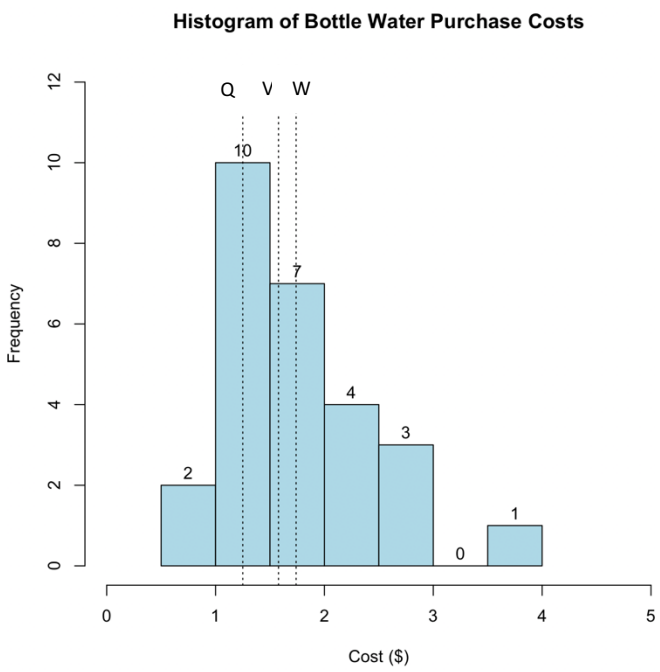
-Directions-

This exam includes 25 multiple-choice questions and three open-response questions that might be used as tie breakers. For questions 1 through 25 (the multiple-choice questions), mark your answer choice in the appropriate location on the sheet provided. After completing questions 1 through 25, answer each tie breaker question in sequential order (i.e., complete Question #1 first, then Question #2, and then Question #3 last). Be sure that your name is printed on each of the tie breaker questions. When time is called, you will be asked to turn in your multiple-choice question answer sheet and your written responses to the tie breaker questions.

1. The graph to the right represents the purchase cost in US dollars ($) for a sample of bottles of water. How many bottles of water are included in the sample? Please disregard the labels Q, V, and W for this question.



   a. 4
   b. 27
   c. 6
   d. Cannot Be Determined

2. Refer to the graph in Question 1. Let Q, V, and W represent measures of center (i.e., central tendency) for the sample data. Provide the correct label for each measure of center below.

   a. $\bar{X} = Q$, Median $= V$, Mode $= W$
   b. $\bar{X} = V$, Median $= W$, Mode $= Q$
   c. $\bar{X} = W$, Median $= V$, Mode $= Q$
   d. Cannot Be Determined

3. Assume for two events, A and B: $P(A) = 0.72$, $P(B) = 0.68$, $P(A \cup B) = 0.80$.
   Do not assume anything further about the events A and B. Find $P(A|B)$.
   Round to 3 decimal places.

   a. 0.720
   b. 0.680
   c. 0.833
   d. 0.882

For Questions 4–6, refer to the setting and statistical output provided below.

Workplace productivity is reduced by 'cyber-loafing', which is a general term to describe using the internet for non-work related activities. Employees frequently engage in 'cyber-loafing', wasting time on the internet to browse non-work related sites such as social media, video streaming, gaming, or travel booking sites. Despite its perception among employees as a trivial issue, businesses are well aware of the loss of productivity 'cyber-loafing' causes.

Arizona State University designed a software to limit 'cyber-loafing'. The software presents on-screen warnings to employees when using sites that may not be work related and also logs the total amount of time spent on non-work related sites each day, making this data available for managers and supervisors.

A study was conducted to determine if implementing this software would lead to decreased 'cyber-loafing'. Daily time wasted on the internet (minutes) was recorded for 43 employees before the implementation (labeled 'Before') and 43 employees after the implementation (labeled 'After'). Assume that the two samples are ***independent*** simple random samples selected from normally distributed populations.

Provided below are sample summary statistics as well as the results of a two-sample independent t-test of means testing the claim that the mean time wasted on the internet before is higher than the mean time wasted on the internet after implementation (i.e., $\mu_1 > \mu_2$), where μ_1 represents the population mean before and μ_2 represents the population mean after implementation.

Sample Statistics:

Sample	n	Mean	Std. dev.
Before	43	77.244493	34.44879
After	43	37.538583	21.78937

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
μ_1 - μ_2	39.70591	6.2160628	70.969448	6.3876301	<0.0001

4. Referring to the setting provided above, which of the following is the alternative hypothesis?

   a. The mean time wasted on the internet before the implementation is significantly different than the mean time wasted on the internet after the implementation.

   b. The mean time wasted on the internet before the implementation is not significantly different than the mean time wasted on the internet after the implementation.

   c. The mean time wasted on the internet before the implementation is significantly higher than the mean time wasted on the internet after the implementation.

   d. The mean time wasted on the internet before the implementation is not significantly higher than the mean time wasted on the internet after the implementation.

5. Refer to the statistical output. Using a 0.05 significance level, which of the following is the most appropriate conclusion for the hypothesis test given the results?

    a. Reject the null hypothesis; there is sufficient evidence to suggest that the mean time wasted on the internet before the implementation is significantly higher than the mean time wasted on the internet after the implementation.

    b. Reject the null hypothesis; there is not sufficient evidence to suggest that the mean time wasted on the internet before the implementation is significantly higher than the mean time wasted on the internet after the implementation.

    c. Reject the null hypothesis; there is sufficient evidence to suggest that the mean time wasted on the internet before the implementation is significantly different than the mean time wasted on the internet after the implementation.

    d. Accept the null hypothesis; there is not sufficient evidence to suggest that the mean time wasted on the internet before the implementation is significantly higher than the mean time wasted on the internet after the implementation.

6. Given the goal of the study (i.e., the pre-post mean comparison), would utilizing a matched pairs design better the experimental design? That is, would utilizing a matched pairs design be ***more*** appropriate for determining the effect software implementation has on 'cyber-loafing' in the workplace? Please select the best answer of those provided below.

    a. Yes. A matched pairs design would allow for more replicates.

    b. Yes. A matched pairs design would utilize the same employees both before and after the implementation, allowing for better control of extraneous variables.

    c. No. A matched pairs design would reduce the number of replicates.

    d. No. The current sampling procedure (i.e., simple random sampling) and the current experimental design with independent samples is sufficient.

7. Consider Type I and Type II Errors in a pregnancy test setting where the null hypothesis states that a woman is not pregnant. Also, consider early medical intervention beneficial and (in fact) critical for the health of both mother and child. Given this information, which of the following significance levels is preferable in this pregnancy test setting?

    a. 0.10

    b. 0.05

    c. 0.01

    d. Cannot Be Determined

8. Using a representative sample of high school students, a 99% confidence interval (CI) was constructed for the population mean amount of time spent on homework daily (minutes). Given that the constructed 99% CI was (55.453, 83.447), which of the following is false?

    a. The null hypothesis would not be rejected at the 1% significance level when testing the following: $H_0: \mu = 60$ vs. $H_1: \mu \neq 60$

    b. The sample mean amount of time spent on homework daily is 69.450 minutes

    c. 99% of observed values in the sample for time spent on homework daily fall between 55.453 and 83.447 minutes

    d. Using the same data, a 90% CI for the population mean amount of time spent on homework daily (minutes) would be narrower than the 99% CI.

9. A population contains 5 subgroups, each consisting of homogenous (i.e., similar) individuals. Given that a researcher would like a sample that contains individuals from each of these 5 subgroups, which of the following would be the most preferable sampling procedure?

    I. Convenience Sampling
    II. Simple Random Sampling
    III. Stratified Sampling
    IV. Cluster Sampling

    a. Any of the Above

    b. II only

    c. III only

    d. II, III, and IV only

10. Refer to the discrete probability distribution provided in the table below.

$X = x$	1	2	3	4	5
$P(X = x)$	0.085	0.235	0.450	?	?

    Find the probability that x is greater than or equal to 4. Round to 3 decimal places.

    a. 0.770

    b. 0.230

    c. 0.115

    d. Cannot Be Determined

11. Several individuals miss flights each day at airports around the world. Assume that on any given day 3% of individuals that are sick miss their scheduled flights. Assume also that on any given day, 22% of individuals that are not sick miss their scheduled flights. If 8 out of 100 individuals with scheduled flights are sick on any given day, what is the probability that an individual that misses their scheduled flight is actually sick? Round to 3 decimal places.

    a. 0.012
    b. 0.010
    c. 0.970
    d. 0.002

For Questions 12-15, refer to the table, which summarizes data for a sample of 32 individuals.

	Ability to Perform the 'Floss' Dance	
Age	Not Able (No)	Able (Yes)
≤ 18 years (Young)	6	14
> 18 years (Old)	9	3

12. What is the probability of one randomly selected individual being able to perform the 'Floss' given that the individual is greater than 18 years old? Round to 3 decimal places.

    a. 0.176
    b. 0.094
    c. 0.813
    d. 0.250

13. Assume simple random sampling for the data summarized in the table above.
    Let p_F represent the proportion of individuals that are able to perform the 'Floss'.
    Which of the following statistical procedures would be most appropriate to test the claim that more than 35% of individuals are able to do the 'Floss'?
    Assume that all necessary requirements hold.

    a. One-tailed one-sample t-test of means
    b. One-tailed one-sample z-test of proportions
    c. Two-tailed one-sample z-test of proportions
    d. Two-tailed two-sample z-test of proportions

14. Referring to Question 13, what is the appropriate p-value for testing the claim that more than 35% of individuals are able to do the 'Floss'? Round to 3 decimal places.

    a. 0.016
    b. 0.001
    c. 0.766
    d. 0.032

15. Refer to Questions 13-14. Using a 0.01 significance level, what would be the most appropriate conclusion for the hypothesis test given the results? What type of error ***might*** you be making? Provide both answers, selecting the best answer choice of those provided below.

    a. Reject H_0, Possible Type I Error
    b. Reject H_0, Possible Type II Error
    c. Fail to Reject H_0, Possible Type I Error
    d. Fail to Reject H_0, Possible Type II Error

16. The 'Jeep wave' is an established practice where drivers of Jeeps (especially Jeep Wranglers) wave to each other when passing on roadways. However, not all drivers of Jeeps participate in the 'Jeep wave'. It was reported that 87% of all Jeep drivers give the 'Jeep wave'. Assuming there are 15 Jeep drivers, what is the expected number that give the 'Jeep wave'? Round to the nearest whole number.

    a. 13
    b. 2
    c. 6
    d. 15

17. The minimum of a distribution is 3, the 25^{th} percentile is 5, the interquartile range (IQR) is 4, and the range is 13. Provide the maximum of the distribution and determine if it is an outlier based upon the 1.5xIQR 'rule'. Please select the best answer of those provided below.

    a. The maximum is 16, it is an outlier
    b. The maximum is 16, it is not an outlier
    c. The maximum is 16, whether or not it is an outlier cannot be determined
    d. The maximum and whether or not it is an outlier cannot be determined

18. The time n riders spent waiting for a roller coaster (minutes) as well as the duration of the roller coaster ride (seconds) was recorded. This data was used to construct a simple linear regression equation $\hat{y} = b_0 + b_1 x$, where x represents the duration of the roller coaster ride (seconds) and y represents the wait time to get on the roller coaster (minutes). Summary statistics are provided in the table below.

	Duration (seconds)	Wait Time (minutes)
Mean	112	53
Standard Deviation	51	27

Given the summary statistics in the table, which of the following values could be the slope of the least squares regression line?

a. 3.476

b. 0.227

c. 1.224

d. 0.635

19. Consider two sets of quantitative data (Dataset G and Dataset Q), each with non-zero variance. If every value in Dataset G is shifted by a constant, which of the following sample statistics will be changed?

I. The mean of Dataset G
II. The mode of Dataset G
III. The range of Dataset G
IV. The linear correlation coefficient between Dataset G and Dataset Q

a. All of the Above

b. I and II only

c. I, II, and III

d. I, II, and IV

20. The Elo rating system is a method to calculate relative skill level of players in games such as chess. Elo ratings for adult tournament chess players are normally distributed, with a mean of 1550 and a variance of 46225. A particular individual has an Elo rating with a corresponding z-score of 1.3. What is the Elo rating of this individual? Round to the nearest whole number.

a. 61643

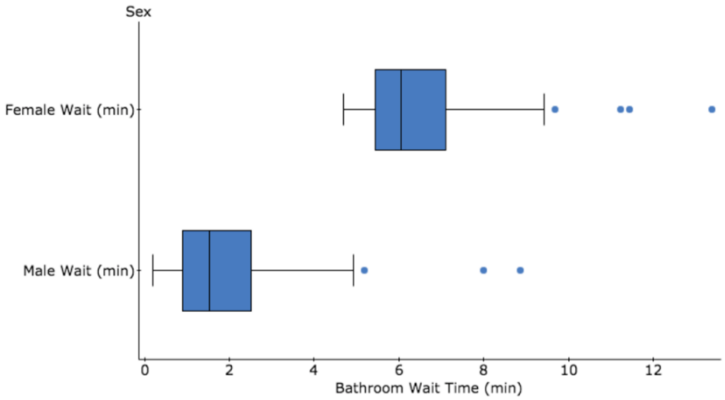
b. 1271

c. 1830

d. 2015

21. Bathroom breaks are not always convenient, especially for females! Refer to the comparison boxplot of bathroom wait times (minutes) for male and female students during a class change.

    Referring to the visual provided, what can we say about bathroom wait times for males and females?



Please select the best answer of those provided below.

a. The distributions for male and female wait times appear to have the same mean and to have similar variances.

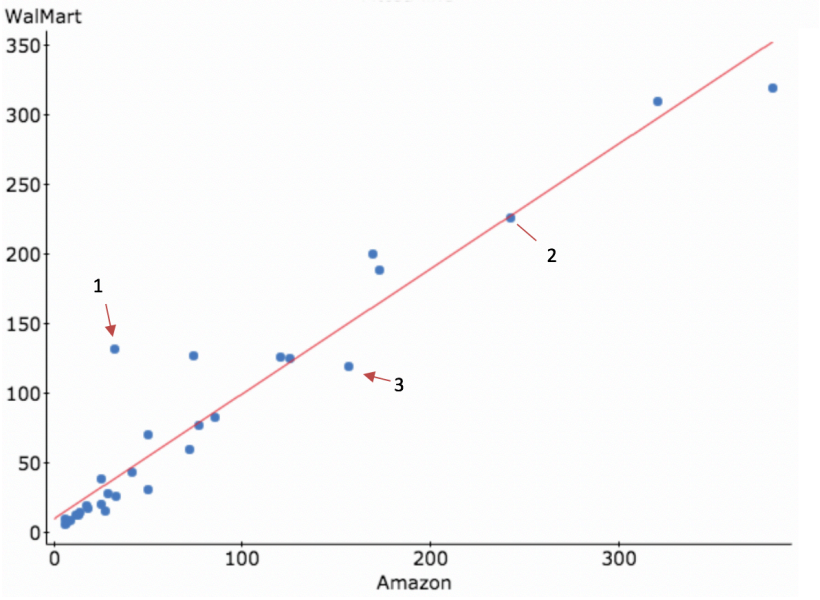
b. The distributions for male and female wait times appear to have the same mean but to have very different variances.

c. The distributions for male and female wait times appear to have very different means but to have similar variances.

d. The distributions for male and female wait times appear to have very different means and to have very different variances.

22. In the scatterplot of Walmart product prices versus Amazon product prices, the fitted simple linear regression line is also shown. Which of the points (1, 2, or 3) has the largest residual? Is this largest residual positive or negative? Provide both answers.

a. 1, Positive
b. 2, Positive
c. 1, Negative
d. 3, Negative



23. A 95% confidence interval for the proportion of currently advertised yoga leggings that have mesh cutouts is calculated using a simple random sample of n currently advertised yoga leggings. Given that the 95% confidence interval is (0.175, 0.418), what is the standard error of the sample proportion (i.e., the standard deviation of the sample proportion)? Please select the best answer of those provided below, rounding to 3 decimal places as necessary.

    a. 0.122
    b. 0.062
    c. Not Enough Information; we would need to know the sample size (n)
    d. Not Enough Information; we would need to know both the sample size (n) and the sample proportion ($\hat{p}$)

24. Assume that the number of typos an individual makes when typing a 5-page essay without proof-reading software is normally distributed with a mean of 21 typos and a standard deviation of 9 typos, what is the probability that a randomly selected sample of 40 individuals typing 5-page essays has a mean number of typos equal to 25 or more? Round to 3 decimal places.

    a. 0.328
    b. 0.002
    c. 0.672
    d. 0.998

25. A dreidel is a four-sided spinning top with a different Hebrew letter on each of the four sides. Each side (i.e., each letter) is equally likely to come up in a single spin of the dreidel. Consider spins independent. When spinning the dreidel 3 times, what is the probability that any one of the four letters comes up all 3 times? Round to 3 decimal places.

    a. 0.016
    b. 0.063
    c. 0.250
    d. 0.422

-Tie Breaker Question 1-

Peter, a university student, finishes his chemistry lab at a random time between 5 and 6 pm. After finishing his lab, he goes directly to the bus station and catches a bus home from campus. It takes him exactly 3 minutes to get to the bus station from his chemistry lab. Two different bus routes (22 Blue and 26 Red) have a bus stop located near his home so he could take either.

Each evening, Peter takes the bus that will depart from the bus station first (either route). His roommate claims that Peter ***always*** takes the 22 Blue route bus, since Peter has taken the 26 Red route bus home only twice in the last 10 weeks. However, Peter claims this is simply a fluke because there should be a 50-50 chance of taking either bus route home each evening.

The departure times from the main bus station are listed below for each bus route.
Assume that the buses are always on time and always depart according to the schedule below.

22 Blue Route	26 Red Route
5:05 pm	5:07 pm
5:15 pm	5:17 pm
5:25 pm	5:27 pm
5:35 pm	5:37 pm
5:45 pm	5:47 pm
5:55 pm	5:57 pm
6:05 pm	6:07 pm

Explain why Peter rarely takes the 26 Red route bus home.
You must provide reasoning for your answer.

-Tie Breaker Question 2-

The table below represents data from a consumer behavior survey and provides the age and the preferred online retailer for a random sample of individuals. The goal of the study was to determine if there was a statistically significant association between the age and the online retailer preference of an individual.

	Age		
Preferred Online Retailer	**18-34**	**35-54**	**55+**
Amazon	7	97	10
Walmart	25	76	23
Target	21	27	24
No Preference	13	6	26

Do the data indicate an association between age and online retailer preference of an individual? Conduct an appropriate hypothesis test to answer this question using a 0.05 significance level. Provide the hypotheses, test statistic(s), p-value(s), and a formal conclusion.

-Tie Breaker Question 3-

A student athlete, mathlete, and bandie is trying to arrange her class schedule for the year. Maria must take a total of 7 specific classes for the 7 hours of the school day. However, band is during 1st hour and the athletic hour that she must register for is during 7th hour.

Counselors at the school will randomly arrange the rest of her schedule, given there are no time constraints on her other courses. All of Maria's other courses are offered during every school hour. Assume that Maria's friend Rose is already registered for AP Calculus AB during 3rd hour.

What is the probability that Maria will have AP Calculus AB during the same hour as her friend Rose? Round to 3 decimal places. You must provide reasoning for your answer.

Multiple Choice Key

1. b
2. c
3. d
4. c
5. a
6. b
7. a
8. c
9. c
10. b
11. a
12. d
13. b
14. a
15. d
16. a
17. a
18. b
19. b
20. c
21. c
22. a
23. b
24. b
25. b

-Tie Breaker Question 1-

ABBREVIATED Question: Peter, a university student, finishes his chemistry lab at a random time between 5 and 6 pm. After finishing his lab, he goes directly to the bus station and catches a bus home from campus. It takes him exactly 3 minutes to get to the bus station from his chemistry lab. Two different bus routes (22 Blue and 26 Red) have a bus stop located near his home so he could take either. Each evening, Peter takes the bus that will depart from the bus station first (either route). His roommate claims that Peter ***always*** takes the 22 Blue route bus, since Peter has taken the 26 Red route bus home only twice in the last 10 weeks. However, Peter claims this is simply a fluke because there should be a 50-50 chance of taking either bus route home each evening.

Explain why Peter takes the 26 Red Route bus home so infrequently.
You must provide reasoning for your answer.

Solution/Reasoning (One Possible Answer):

Notice that the 50-50 chance is simply something Peter believes to be true but is not a stated probability of taking either bus route home.

Looking at the table of departure times, Peter is (in fact) less likely to take the 26 Red route.

Despite Peter getting out of his chemistry class at a random time, the bus routes are fixed and assumed to run on time. Thus, in any 10 minute interval (e.g., from 5:03 pm to 5:13 pm), there are 8 minutes such that Peter would take the 22 Blue route (i.e., from 5:03 pm to 5:05 pm and again from 5:07 pm to 5:13 pm). However, there are only 2 minutes such that Peter would take the 26 Red route (i.e., from 5:05 pm to 5:07 pm). Thus, Peter takes the 26 Red route two times in 10 weeks, which is consistent with his observed bus rides.

Rubric:

0 pts – No Answer Provided or Probability Calculated Based on 50-50 Statement

1 pts – Answer Provided but No Mention of Departure Times, Time Gaps, etc.

2 pts – Answer Provided with Mention of Departure Times but No Further Explanation of Time Gaps

3 pts – Answer Provided with Mention of Departure Times and Departure Time Gaps

4 pts – Answer Provided with Mention of Departure Times and Departure Time Gaps with Example

-Tie Breaker Question 2-

	Age		
Preferred Online Retailer	**18-34**	**35-54**	**55+**
Amazon	7	97	10
Walmart	25	76	23
Target	21	27	24
No Preference	13	6	26

Do the data indicate an association between age and online retailer preference of an individual? Conduct an appropriate hypothesis test to answer this question using a 0.05 significance level. Provide the hypotheses, test statistic(s), p-value(s), and a formal conclusion.

Solution:

- **Hypotheses**
  $\begin{cases} H_0: \text{Age and Online Retailer Preference are Independent (there is no association)} \\ H_1: \text{Age and Online Retailer Preference are Dependent (there is association)} \end{cases}$
- **Test Statistic**
  $\chi^2 = 89.367,\ df = 6$
- **P-Value**
  $p - value < 0.0001 \approx 0$
- **Formal Conclusion at $\alpha = 0.05$ (in terms of H_0)**
  Reject the null hypothesis at the 5% significance level. There is sufficient evidence to support the claim that age and online retailer preference of an individual are associated/dependent.

Rubric: 0 pts to 4 pts Possible

1 point for each completely correct answer and 0 points for each incorrect answer of the following tie breaker components: (1) hypotheses, (2) test statistic, (3) p-value, and (4) formal conclusion.

-Tie Breaker Question 3-

A student athlete, mathlete, and bandie is trying to arrange her class schedule for the year. Maria must take a total of 7 specific classes for the 7 hours of the school day. However, band is during 1st hour and the athletic hour that she must register for is during 7th hour.

Counselors at the school will randomly arrange the rest of her schedule, given there are no time constraints on her other courses. All of Maria's other courses are offered during every school hour. Assume that Maria's friend Rose is already registered for AP Calculus AB during 3rd hour.

What is the probability that Maria will have AP Calculus AB during the same hour as her friend Rose? Round to 3 decimal places. You must provide reasoning for your answer.

Solution: $P(\text{Maria has AP Calculus AB during } 3^{rd} \text{ hour}) \approx 0.200$

**** Steps***

1. Notice that of Maria's 7 courses, 2 courses already have fixed time slots. Thus, only 5 courses will be randomly arranged during the day.

The # of ways to arrange 5 courses in 5 hours is $5! = 120$ ways

2. Notice that since Rose already has AP Calculus AB during 3rd hour, the only way Maria could have the course during the same hour as her friend is if she also is assigned to take AP Calculus AB during 3rd hour. Think of fixing her schedule such that this event occurs. Thus, only 4 courses will need to be randomly arranged during the day. This represents the number of ways to arrange 5 courses if the 3rd hour **must** be AP Calculus AB.

The # of ways to arrange 4 courses in the remaining 4 hours is $4! = 24$ ways

3. Given the above, the probability that Maria will have AP Calculus AB during the same hour as her friend Rose is equivalent to the following:

$$P(\text{Maria has AP Calculus AB during } 3^{rd} \text{ hour}) = \frac{\text{\# of ways to arrange 4 remaining courses}}{\text{Total \# of ways to arrange 5 courses}}$$

$$= \frac{24}{120} \approx 0.200$$

Rubric: 0 pts to 4 pts Possible

1 point for each correctly calculated count of the following: (1) # of ways to arrange 5 courses and (2) # of ways to arrange 4 non-fixed courses

If the correct final solution is provided, 4 points should be awarded as long as sufficient work was shown or sufficient reasoning was provided for the answer.